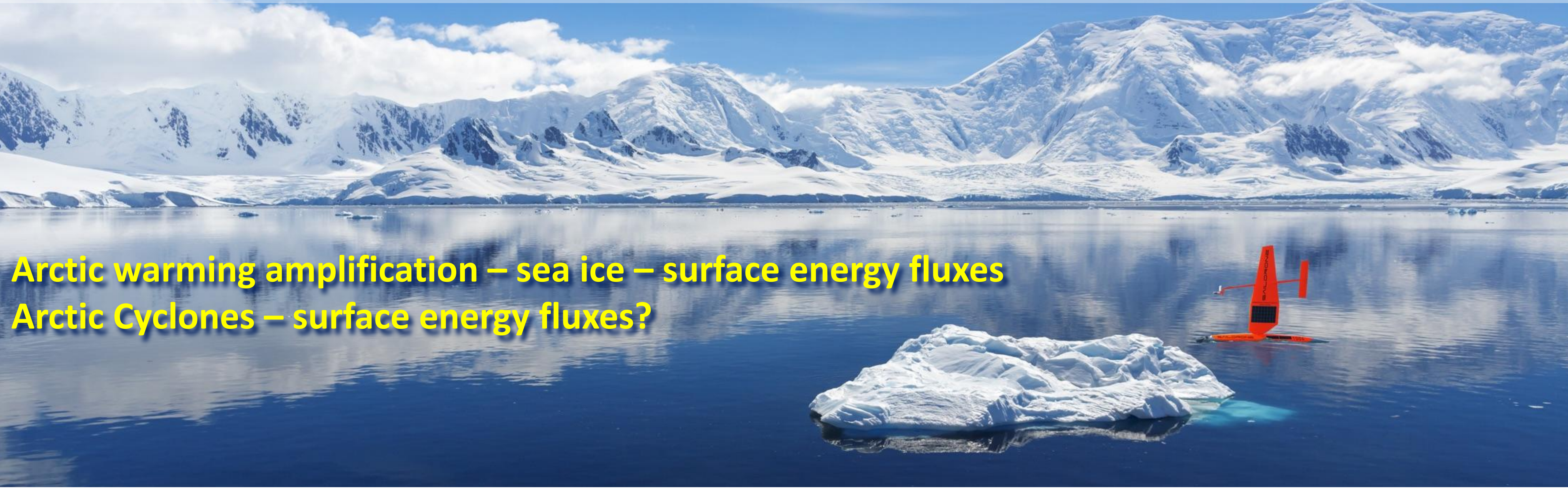
Using Saildrone Observations to Validate Ensemble Forecasts in the Arctic

Chidong Zhang, NOAA Pacific Marine Environmental Laboratory (chidong.zhang@noaa.gov)
Mark Yamane, Subhatra Sivam, and Isabella Dressel, NOAA Hollings Scholarship Program

Nineth NOAA Ensemble Users Workshop
August 22 – 24, 2023
NOAA Center for Weather and Climate Prediction

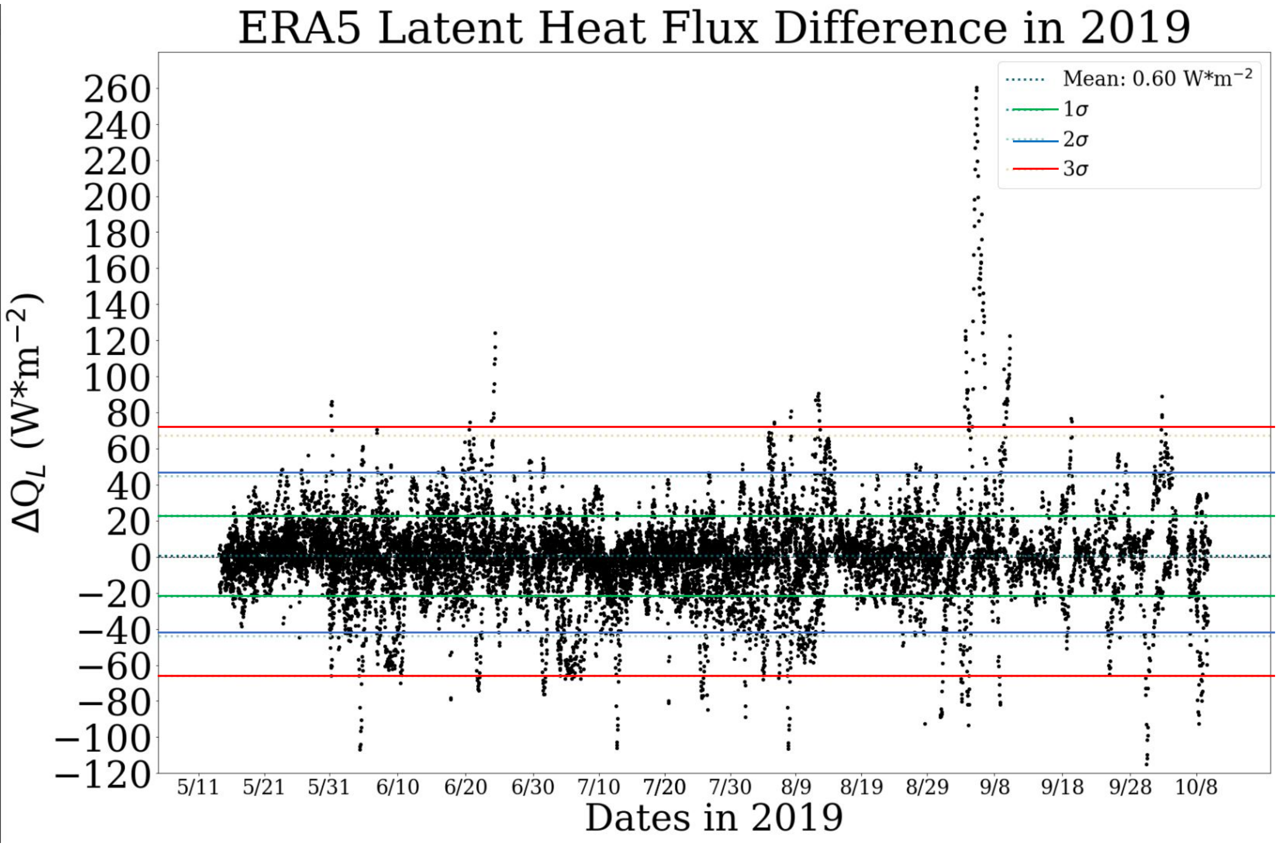
Using Saildrone Observations to Validate Ensemble Forecasts in the Arctic

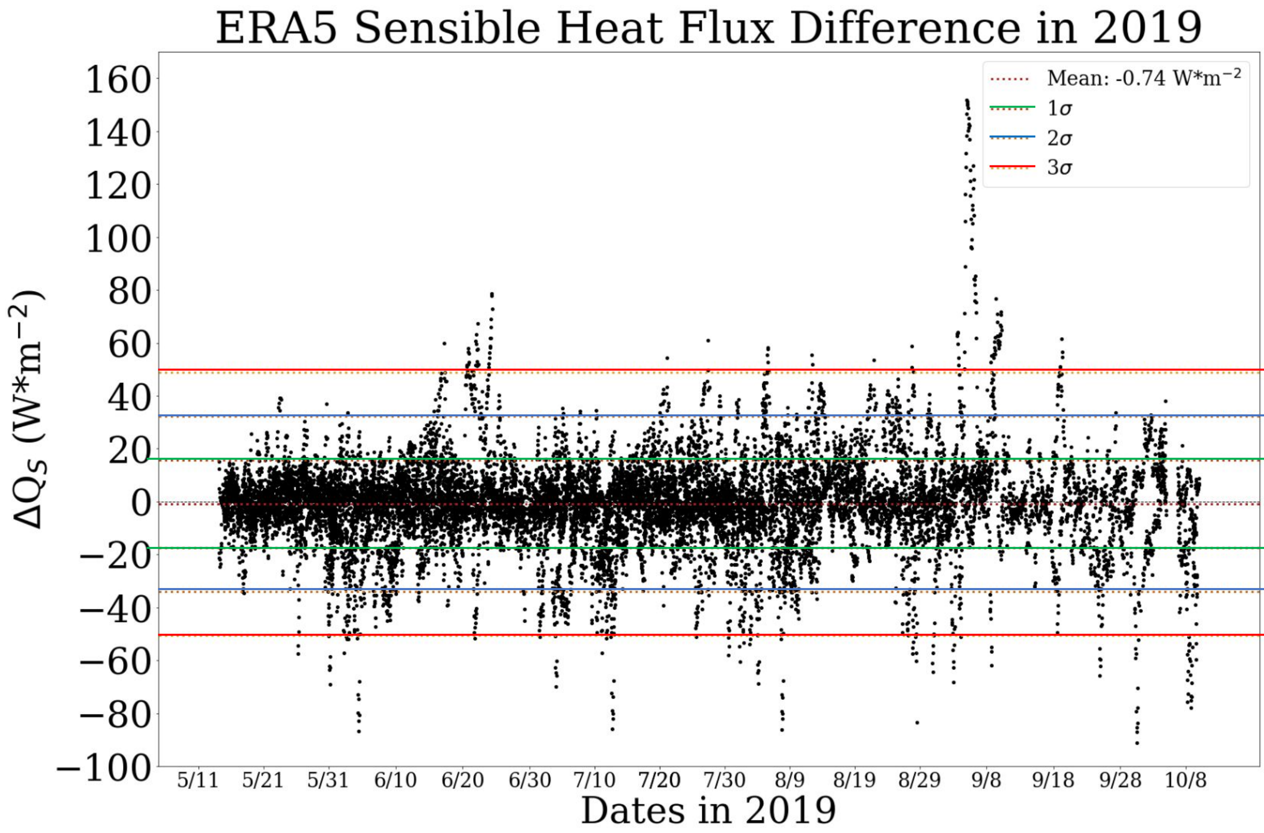
Chidong Zhang, NOAA Pacific Marine Environmental Laboratory (chidong.zhang@noaa.gov)
Mark Yamane, Subhatra Sivam, and Isabella Dressel, NOAA Hollings Scholarship Program



Nineth NOAA Ensemble Users Workshop
August 22 – 24, 2023
NOAA Center for Weather and Climate Prediction

Why Bother? Can't we just use global reanalysis products to validate prediction?





Atmospheric Measurements

1	WIND	Gill Windmaster 3D 20Hz @ + 5.0 m
2	ATRH	Rotronic HC2 - S3 @ + 2.2 m
3	PRESSURE	Vaisala BAROCAP PTB210 @ + 0.2 m
4	RADIATION	LICOR LI-192SA @ + 2.2 m

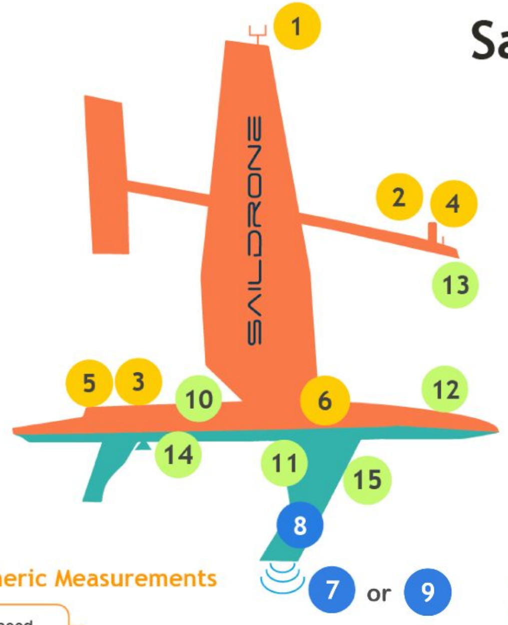
Ocean Measurements

5	CTD	SBE 37 & RBR Conductivity @ -0.5 m
6	DO & TEMP	RBR Coda ODO & SBE 37 ODO @ -0.5m
7	CHL-A	Wetlabs ECO-FL-S G4 & Turner Cyclops-7F
8	SKIN SST	Heitronics CT15.2 @ +2.2 m
9	WAVES	Dual GPS aided IMU - VN 300
10	CAMERAS	Sky, Sea and Horizon Cameras

Options

11	ADCP	Teledyne RDI Workhorse 300 kHz - 2.0 m
	OR	
12	ECHO-SOUNDER	SIMRAD WBT Mini (EK80) @ -2.0 m

Saildrone Sensor Suite



Specifications

Length: **7 m**
Height: **4.6 m (above water line)**
Depth: **2 m**
Weight: **545 kg, (fully loaded)**
Speed: **Transit - 3 Kt, Max - 8 Kt**
Payload Power: **30W Steady state**
Payload Capacity: **100 kg**
Max deployed duration: **12 months**
Longest voyage: **16,100 km**

Atmospheric Measurements

Measurement	#	Sensor
Wind Speed; Wind Direction	1	Anemometer @ +5.0m Gill WindMaster 3D Ultrasonic 20Hz
Sunlight & Infrared Radiation	2	Sunshine Pyranometer @ +2.5m Delta-T Devices SPN1
	3	Pyrgeometer +0.7m Eppley PIR
Air Temperature; Humidity	4	Meteorlogical Probe @+2.4m Rotronic HC2 - S3 with rad shield
		+ Sky and side view cameras
Air Pressure	5	Digital Barometer @ +0.3m Vaisala BAROCAP® PTB210
Air pCO_2	6	CO_2 System @ +0.5m PMEL ASVCO$_2$

Oceanic Subsurface Measurements

Measurement	#	Sensor
Ocean Current	7	ADCP @ -1.8m Teledyne RDI 300 kHz Workhorse Sentinel
Water Temperature	8	RBR or SBE thermistors every 30cm from -0.3m to -1.8m
Fish Biomass; Bathymetry	9	Scientific Echosounder @ -1.8m SIMRAD WMINI Multi-beam Sonar @ -1.8m Norbit iWBMS

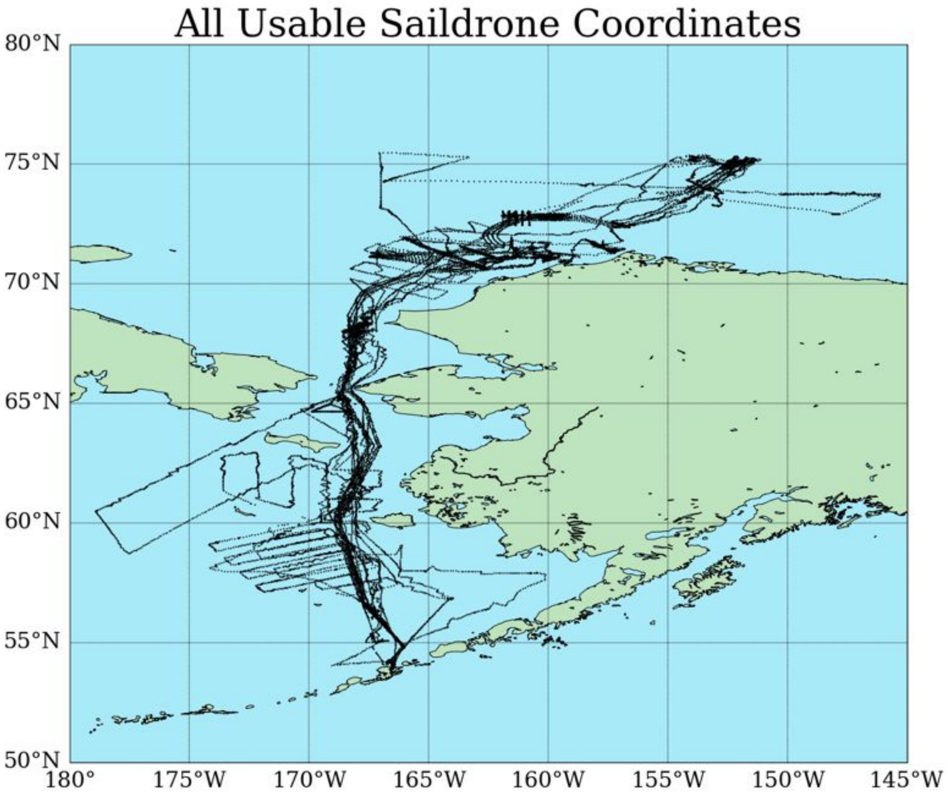
Oceanic Surface Measurements

Measurement	#	Sensor
Wave Height & Period	10	Dual GPS & IMU Vectornav / KVH
Seawater pCO_2 & pH; Dissolved Oxygen; Water Temperature; Salinity	11	CO_2 System PMEL ASVCO$_2$ @ -0.5m Honeywell Durafet @ -0.5m Aanderaa Optode @ -0.5m Sea-Bird Scientific SBE PRAWLER @ -0.6m
Magnetic Field	12	Magnetometer Barrington MAG 648
Skin Temperature	13	SST IR Pyrometer @ +2.2m Heitronics KT15 II
Chla; CDOM Concentration; Red Backscatter	14	Fluorometer and Backscatter @ -0.2m Sea-Bird Scientific WET Labs Eco Triplet
Water Temperature; Salinity	15	Thermosalinograph CTD @ -0.6m SBE37 & RBR conductivity

Question: Can observations from uncrewed surface vehicles (USVs) be used to validate weather forecast?

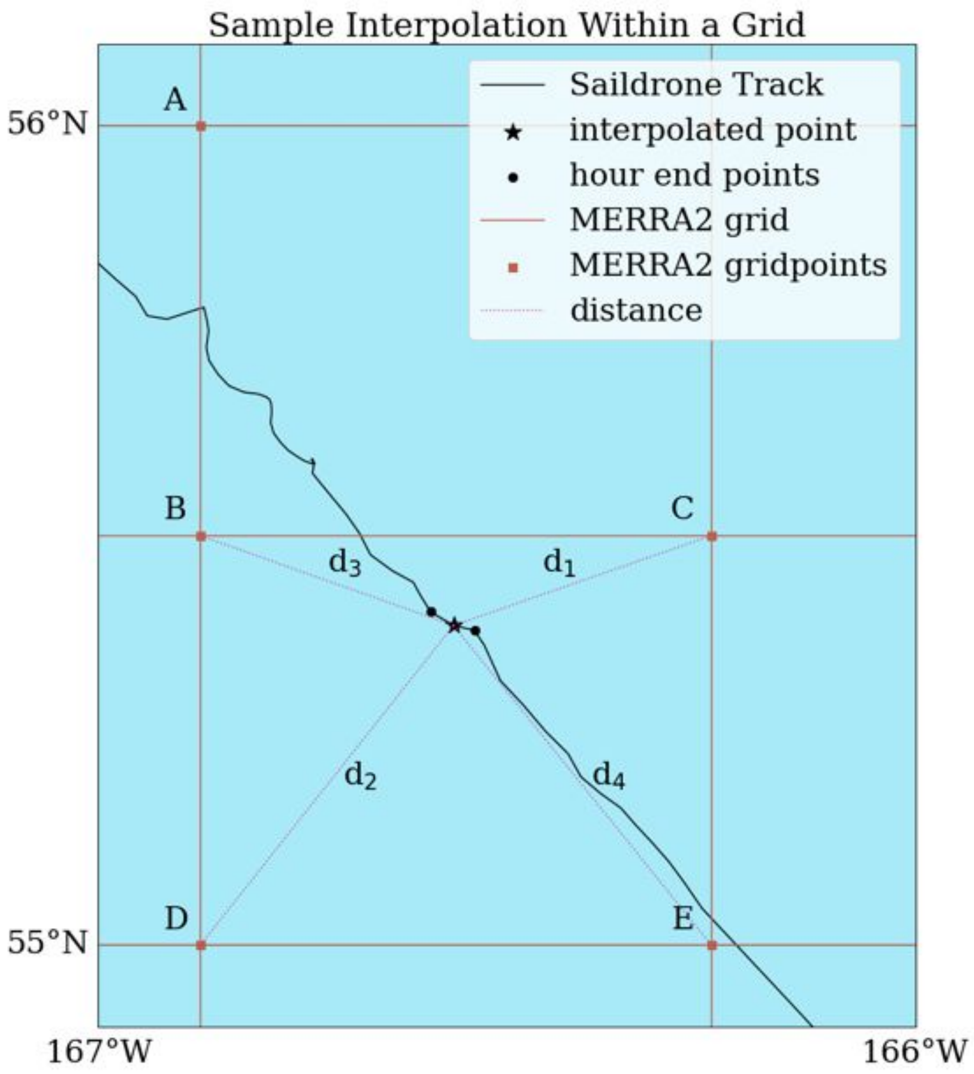
Advantages of USV observations: In regions without other in situ observations (e.g., Arctic oceans, insight hurricanes)

Challenges of using USV observations:
- Moving platforms vs. fixed model grids
- Continuous sampling vs. discrete model output
- Geographical and seasonal coverages vs. sparse sampling
- Ice vs. no ice



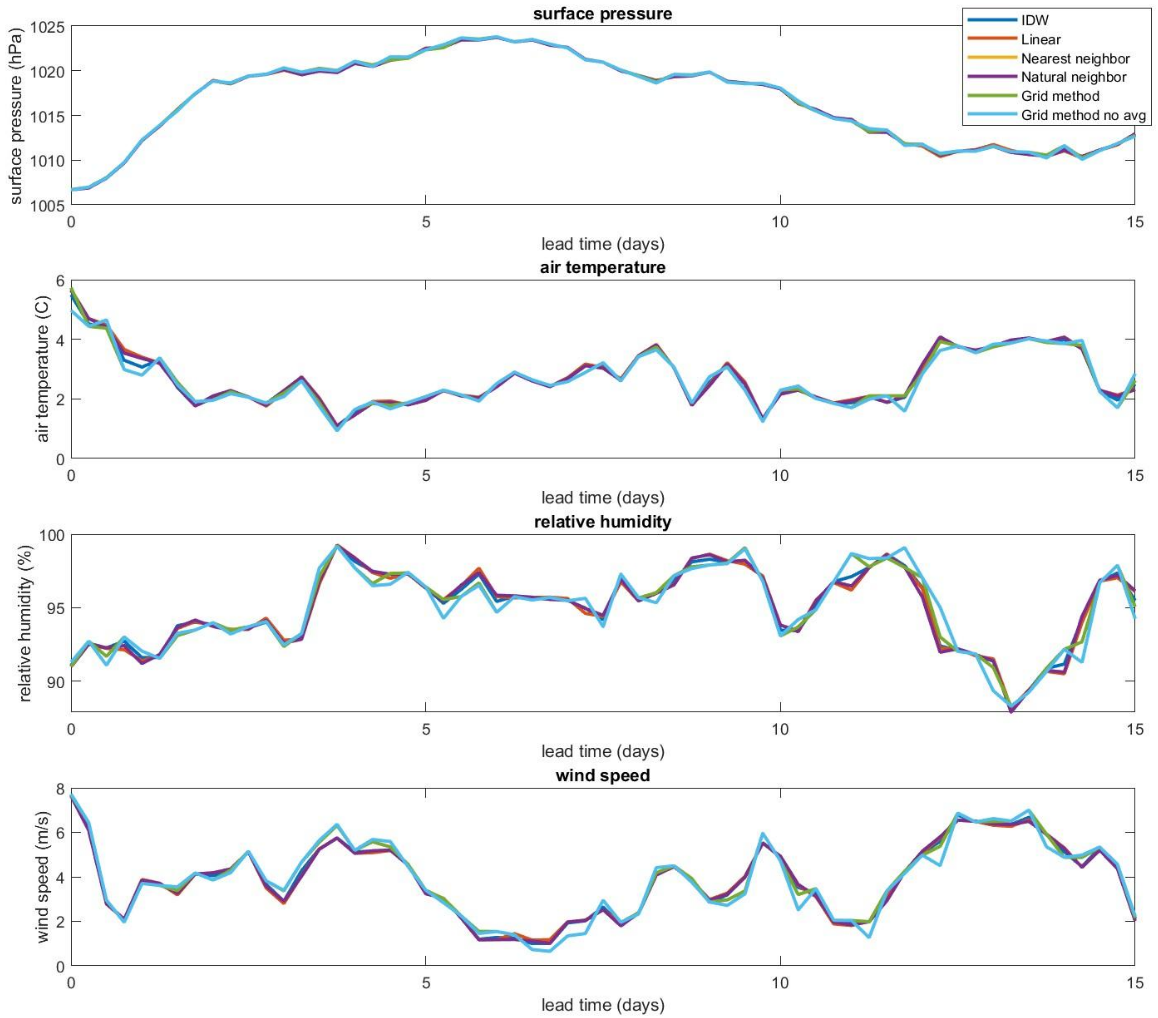
Sivam et al. (2023)

Moving platforms vs. fixed model grids



Different Interpolation Methods

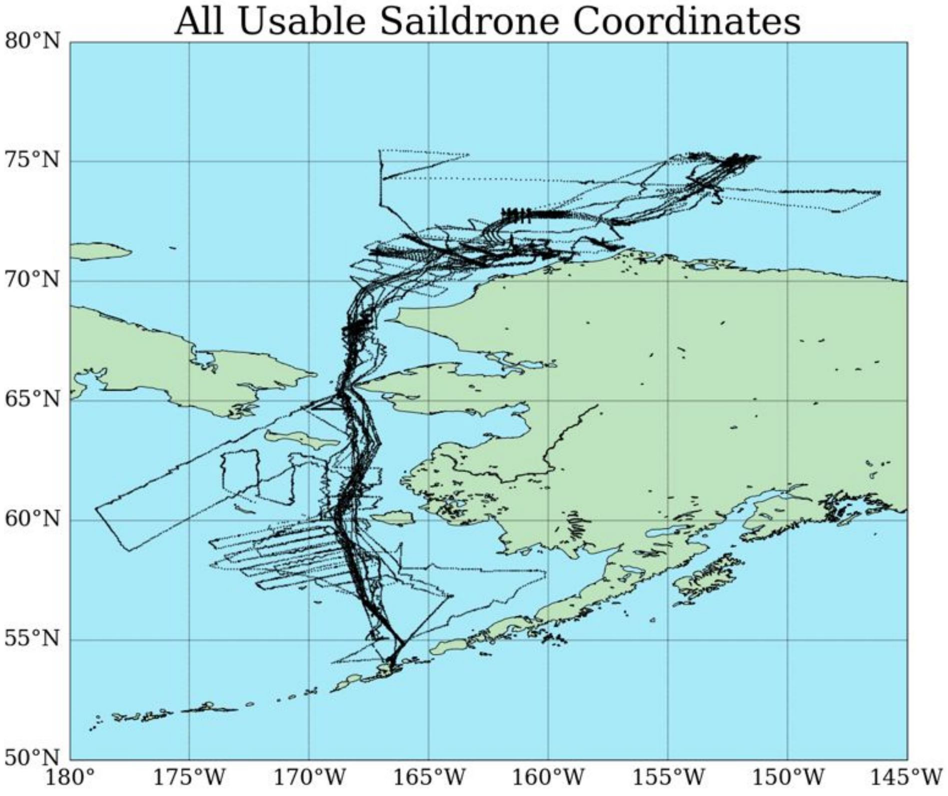
- nearest neighbor
- natural neighbor
- bi-linear linear
- inverse distance weighting

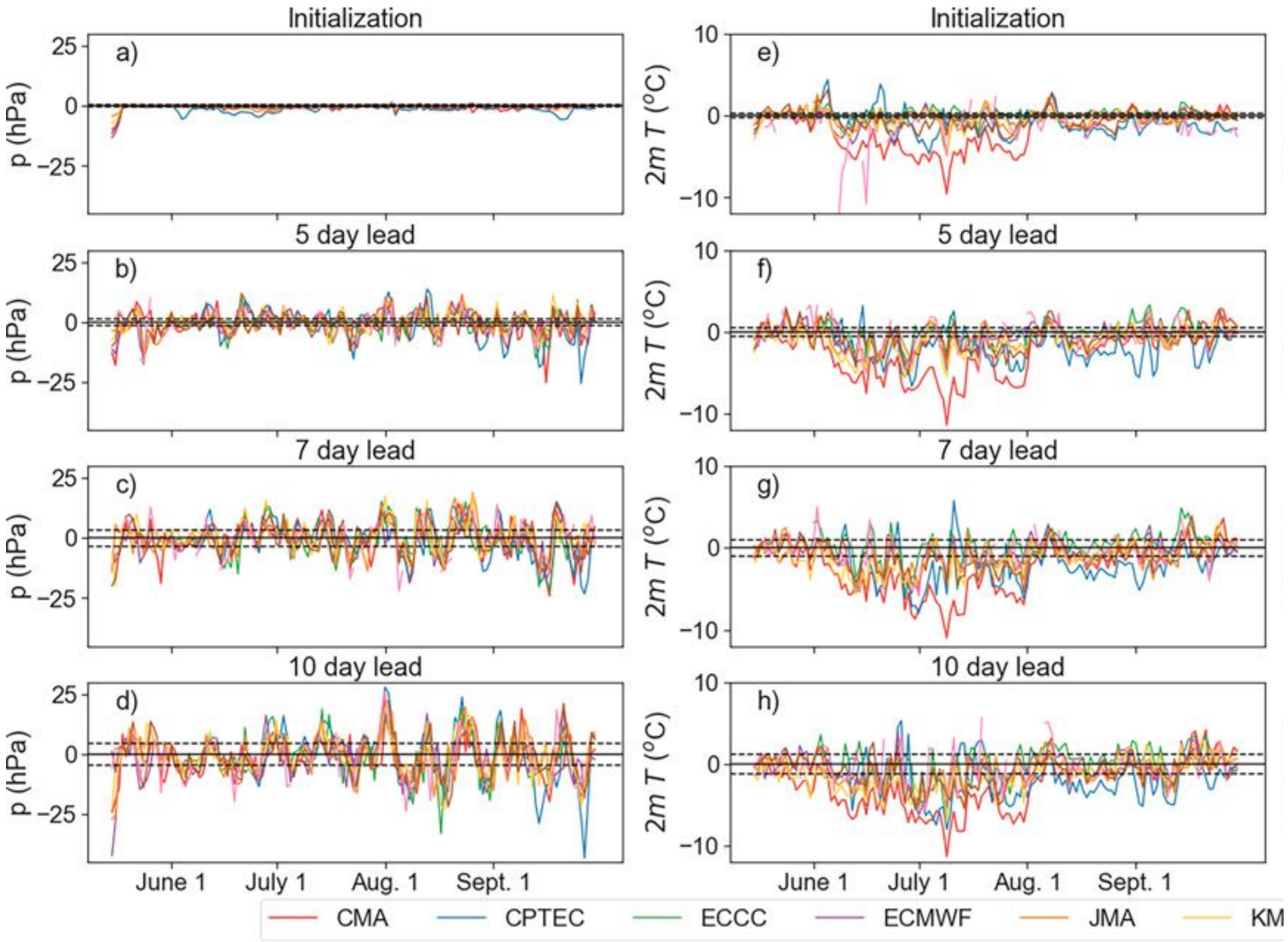


Ice vs. no ice

Geographical and seasonal coverages vs. sparse sampling

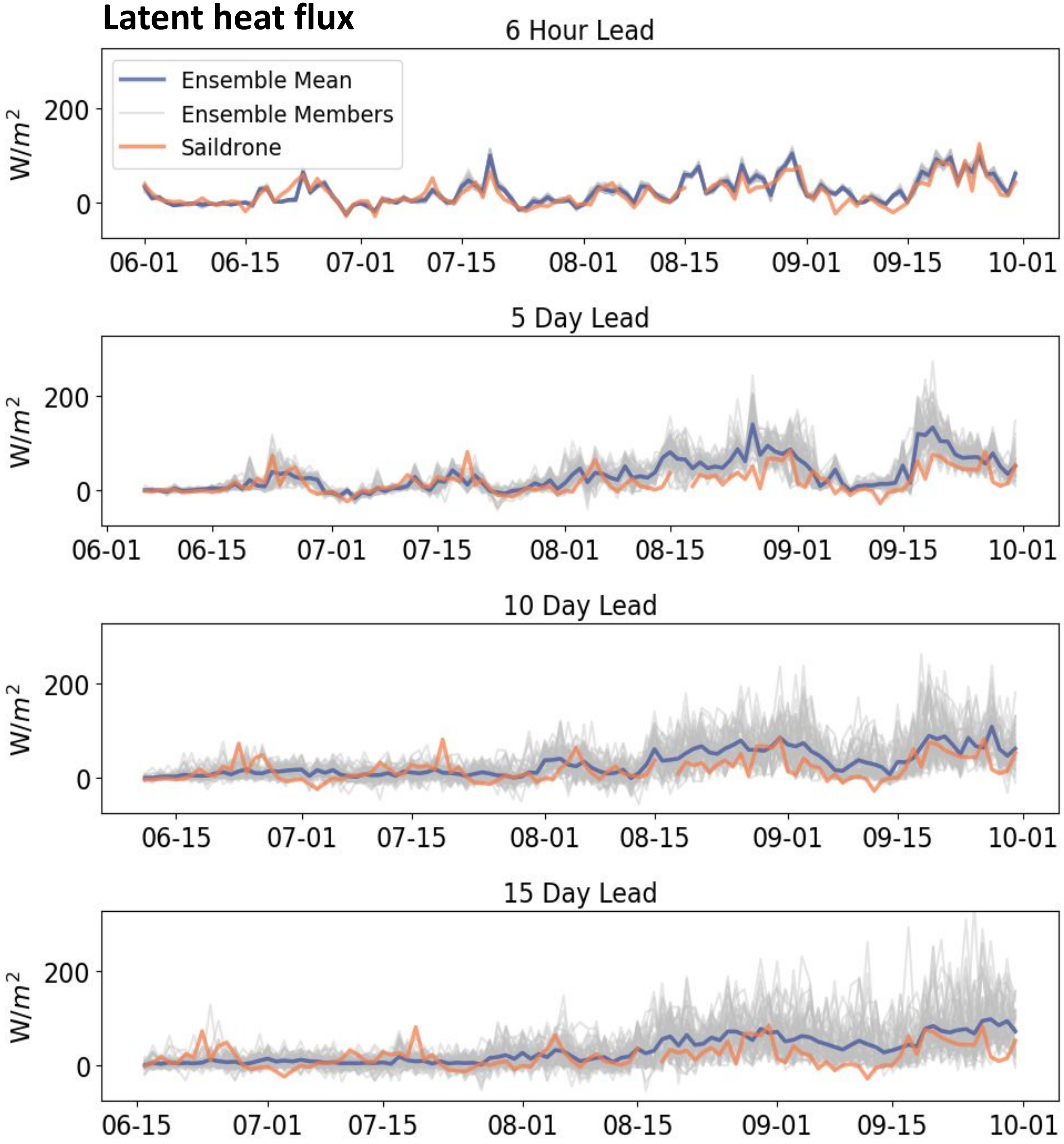
Assumption: The state variables may be geographically and seasonally dependent, their forecast error are not.



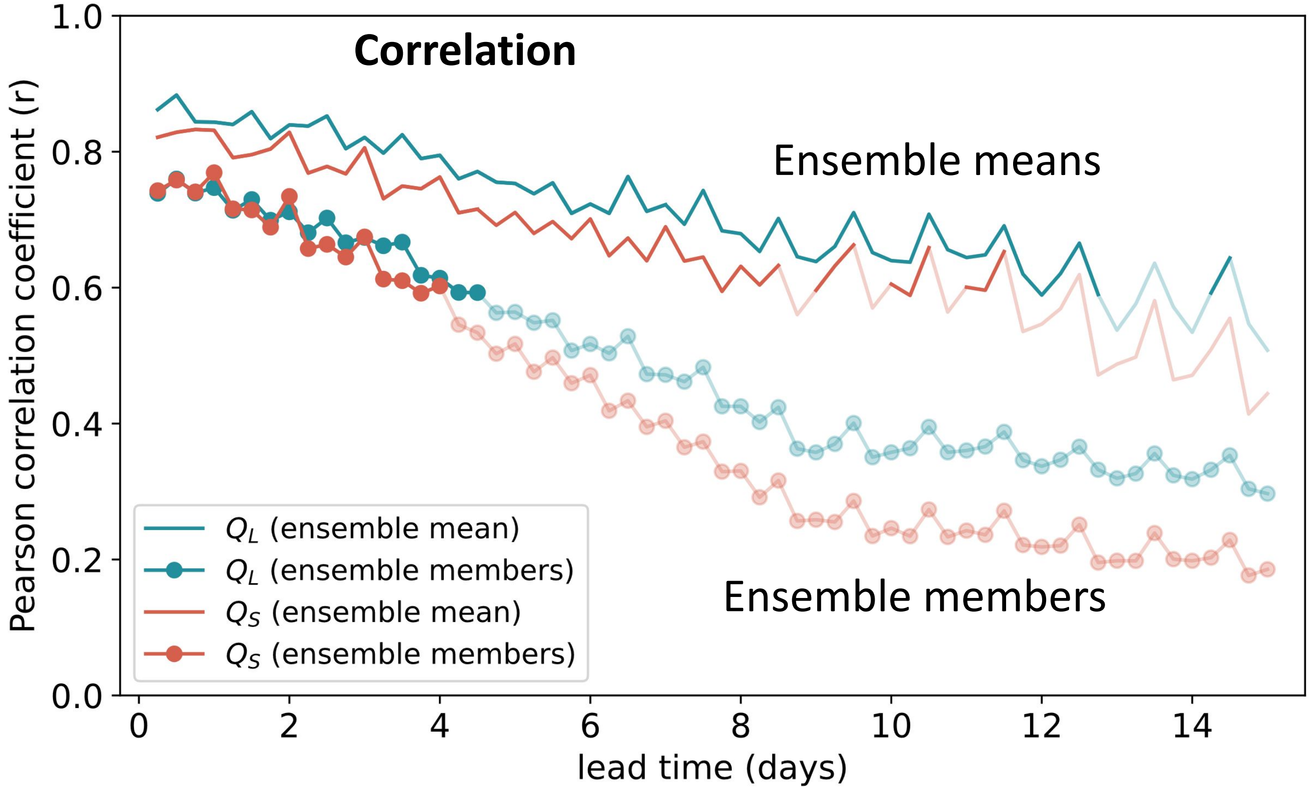


Zhang et al. (2022)

Latent heat flux

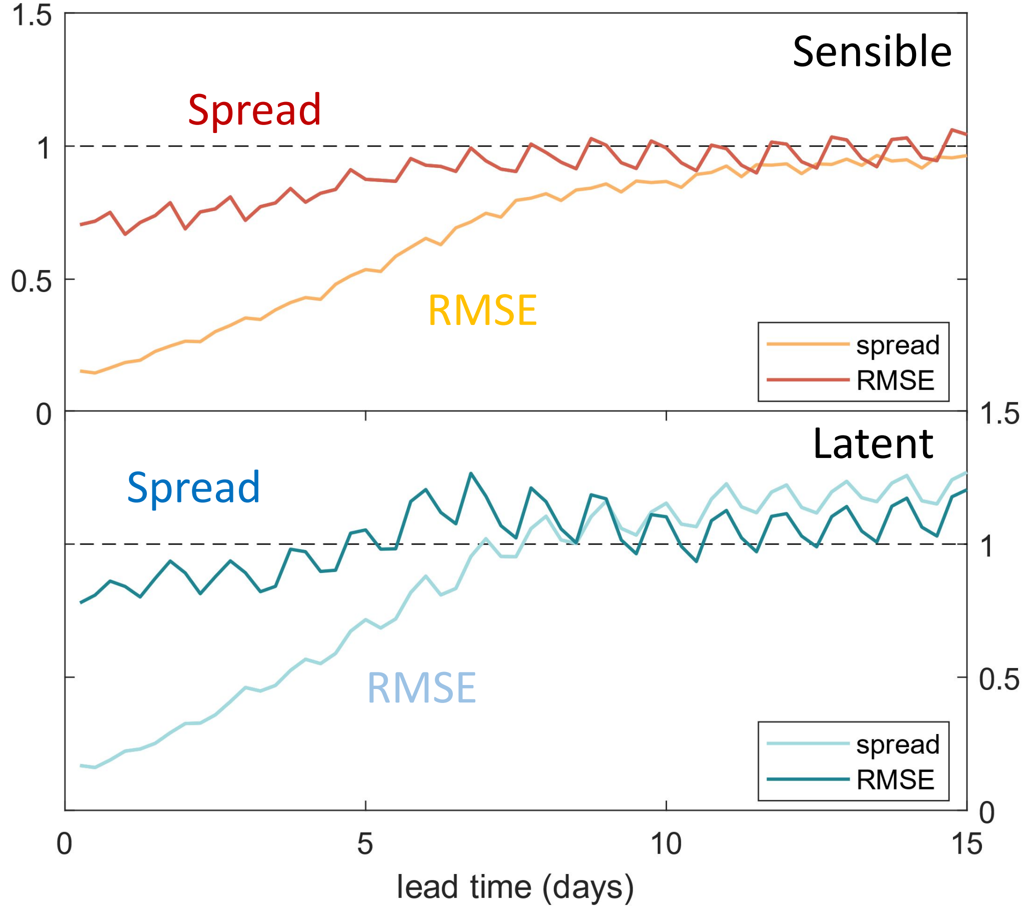


Flux forecast prediction skill and model–observation correlations decline around 4-7 days



Pearson correlation coefficients (r) between ensemble mean and observations as a function of model lead time. Significant (**solid line**) and insignificant (**transparent**) r values.*

*Thresholds for significance are based on the effective sample size (latent: 11, sensible: 15) determined by saildrone autocorrelation using a 95% significance level.

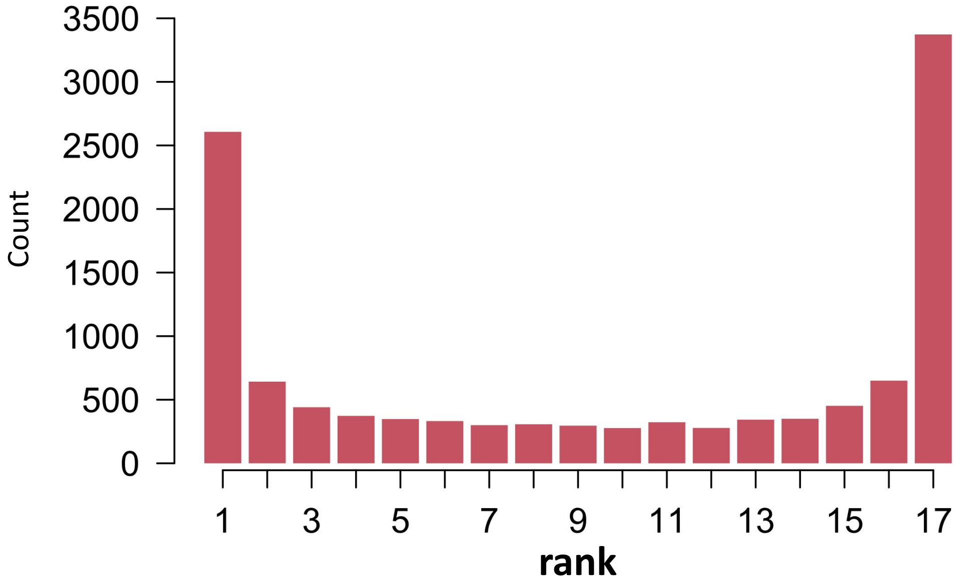


Spread-skill relationship for sensible (top) and latent (bottom) heat fluxes using ensemble spread and RMSE normalized by the observation standard deviation.

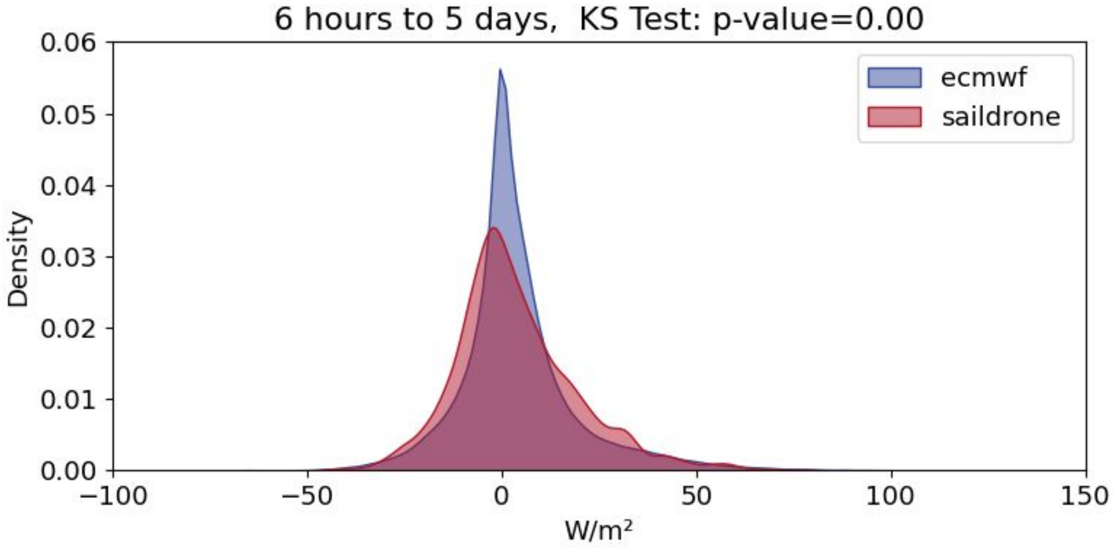
Flux forecasts are underdispersed and overconfident

Sensible heat flux (observation ranks)

(a) 6 hours to 5 days



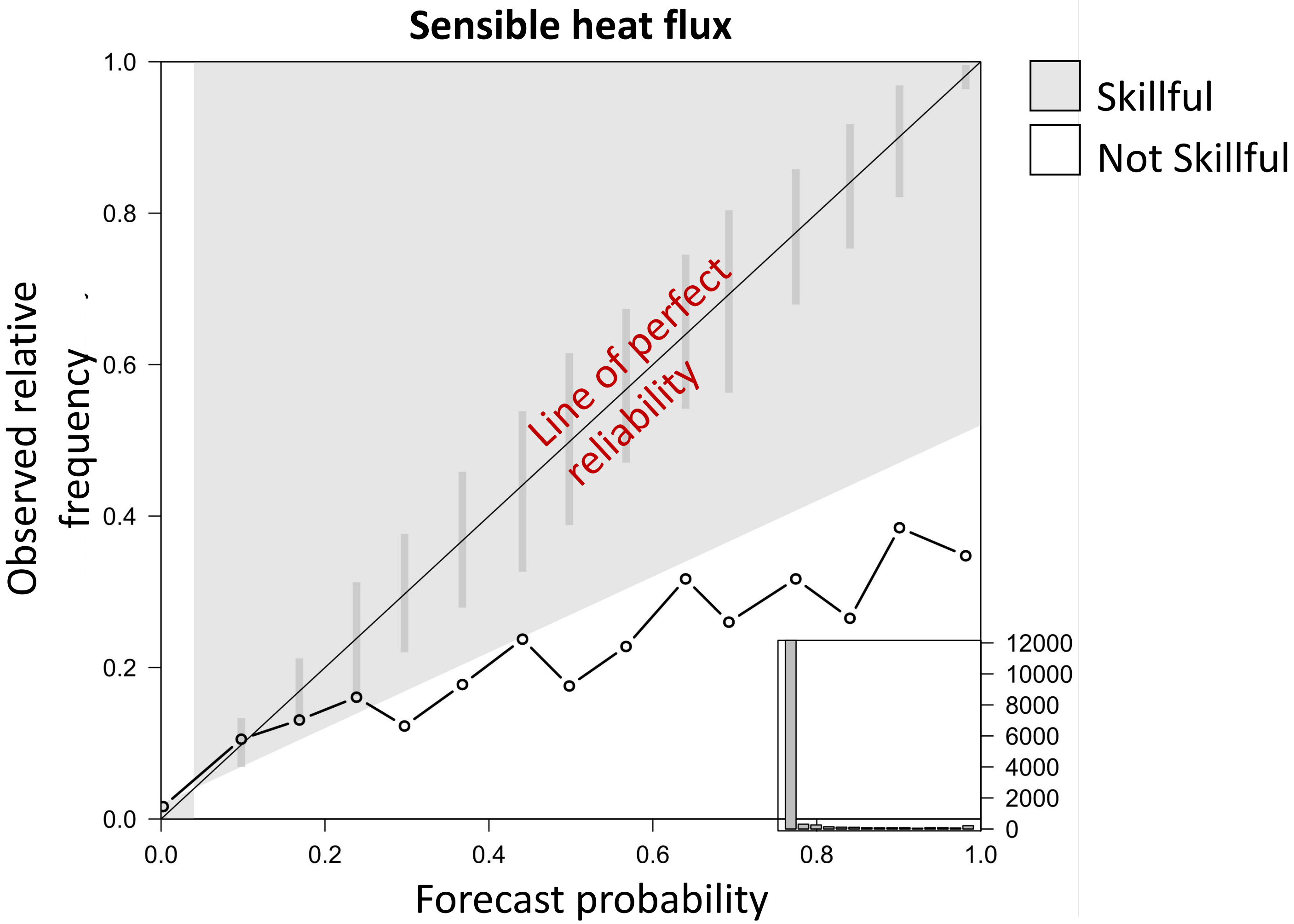
Sensible heat flux (PDFs)



Extreme flux events ($Q_s > 2\sigma$) have unreliable forecasts

Binary probability forecast:
- Within 2σ of mean (0)
- Outside of 2σ of mean (1)

Forecast has **poor resolution** and is and **unreliable** for predicting the extremes



Reliability diagrams for 6-hour to 5-day lead times for sensible heat flux. Probabilities are based on fluxes greater than 2 standard deviations away from the mean.

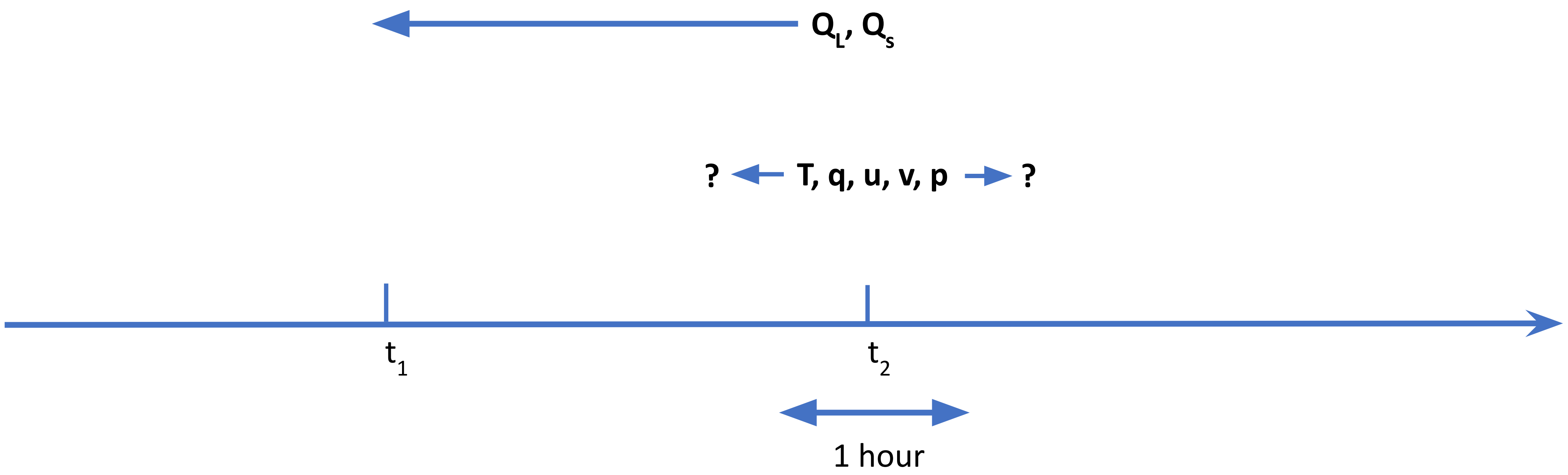
Questions

- How to define ”flux events” for probabilistic forecast validation?
- How to choose validation metrics that may lead to physically meaningful insights?
- How to test the statistical significance of differences validation scores?
- Does it make sense to compare scores for different variables (e.g., fluxes, T, q, V)?

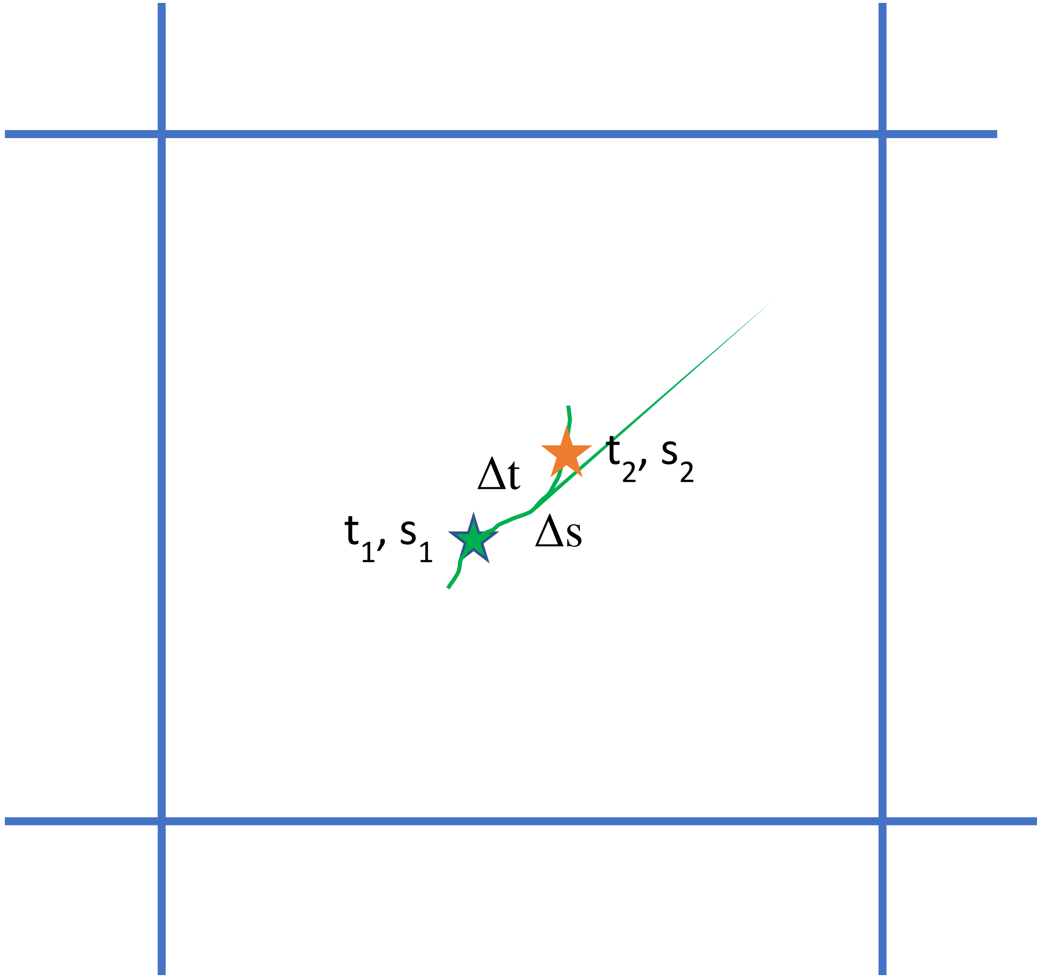
Future Work

- Include more saildrone observations (2017, 2018)
- Include other ensemble forecasts
- Include other validation metrics
- Trace the causes of errors in flux prediction

Continuous sampling of saildrones (every minute) vs. discrete model output (6 hourly)



Error growth due to time and space increments



Let the forecast error be

$$E(s, t) = P(s, t) - O(s, t), \tag{A1}$$

where s represents location, t is time, P is the forecast, and O is the observations. The error increment at a fixed location s_2 over a time period $\delta t = t_2 - t_1$ is

$$\delta_t E(s_2, \delta t) = E(s_2, t_2) - E(s_2, t_1). \tag{A2}$$

For a mobile platform, from time t_1 to t_2, its location would change from s_1 to s_2. So the error increment from time t_1 to t_2 measured at location s_2 is

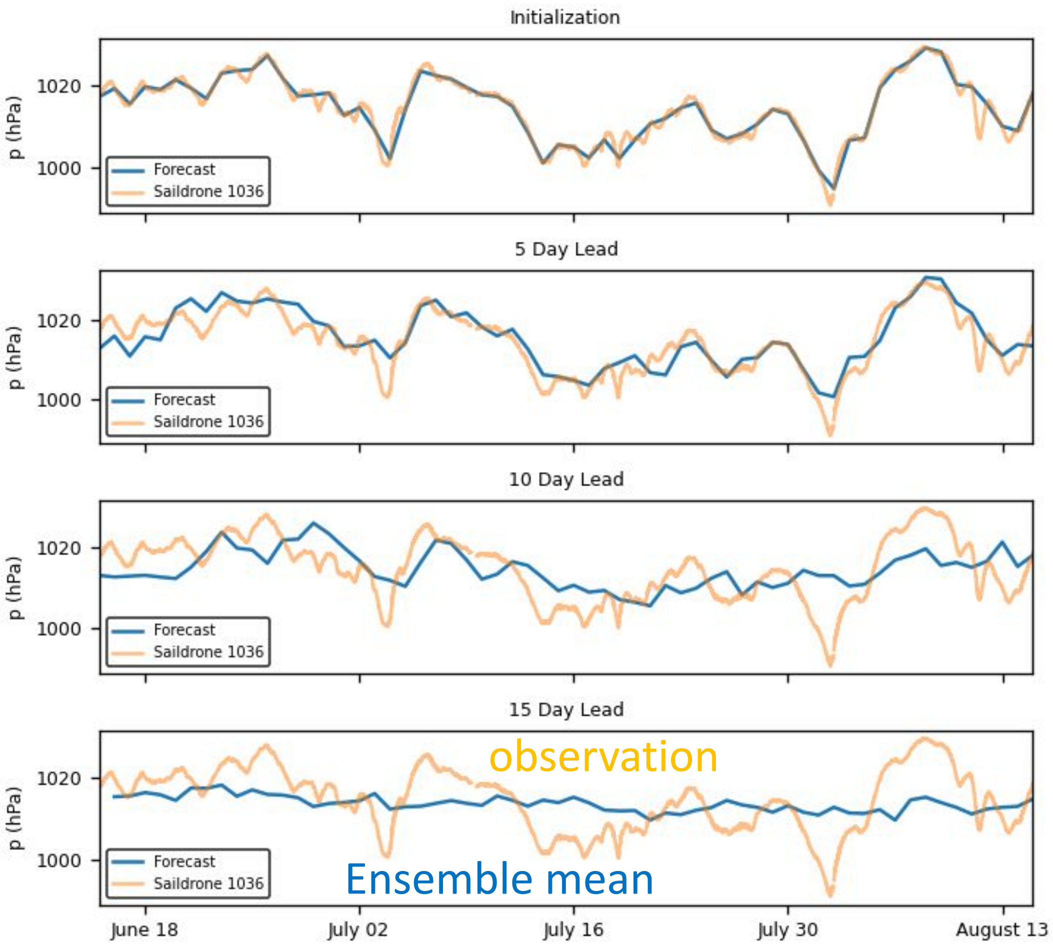
$$\begin{aligned} \delta E(s_2, \delta t) &= E(s_2, t_2) - E(s_1, t_1) \\ &= [E(s_2, t_2) - E(s_2, t_1)] + [E(s_2, t_1) - E(s_1, t_1)] \\ &= \boxed{\delta_t E(s_2, \delta t)} + \boxed{\delta_s E(\delta s, t_1)}, \end{aligned} \tag{A3}$$

where $\delta_t E(s_2, \delta t)$ is the intended measure of the error increment in time [Eq. (A2)], and $\delta_s E(\delta s, t_1) = E(s_2, t_1) - E(s_1, t_1)$ measures the spatial variability at t_1 over the distance $\delta s = s_2 - s_1$ traveled by the mobile platform during δt. The practically measured error increment $\delta E(s_2, \delta t)$ can be taken as a close approximation of the intended error increment in time $\delta_t E(s_2, \delta t)$ only if $\delta_s E(\delta s, t_1)$ is negligibly small.

Zhang et al. (2022)

Examples of ECMWF Ensemble Means

Surface Pressure



Surface Relative Humidity